Sex differences in depressive symptoms and their networks in a treatment-seeking population - a cross-sectional study

Vetter, Johannes Simon ; Spiller, Tobias Raphael ; Cathomas, Flurin ; Robinaugh, Donald ; Brühl, Annette ; Boeker, Heinz ; Seifritz, Erich ; Kleim, Birgit

Abstract: BACKGROUND: The higher prevalence of major depressive disorder (MDD) in females relative to males is well-established. Some authors have posited this difference arises to divergent symptom profiles in females vs. males. However, empirical tests of this hypothesis have yielded equivocal results. Here, we investigate sex differences in MDD of individual symptoms and symptom networks in a treatment-seeking sample. METHODS: We assessed depressive symptoms using Hamilton Depression Rating Scale (HDRS-17) in 590 treatment-seeking adults with MDD (300 females). We examined group differences in symptom endorsement. We investigated symptom networks and estimated Gaussian Graphical Models. Finally, we compared the female and male networks using the Network Comparison Test. RESULTS: Females scored significantly higher in psychological anxiety ($p < 0.001$; rB = -0.155), somatic anxiety ($p = .001$; rB = -0.150) and feelings of guilt ($p = .002$; rB = -0.139). Male and female patients did not differ in depression sum scores. There were no sex differences in network structure or global strength. LIMITATIONS: Our study was sufficiently powered to detect only medium sized symptom differences. The generalizability of our study is limited to clinical samples and further studies are needed to investigate if findings also translate to outpatient samples. CONCLUSION: Females reported elevated anxiety symptoms and guilt. Clinicians should assess these symptom differences and tailor treatment to individual symptom profiles. No differences between sexes emerged in MDD network structures, indicating that features may be more similar than previously assumed. Sex differences in psychopathological features of MDD are important for future research and personalized treatment.

DOI: https://doi.org/10.1016/j.jad.2020.08.074

networks in a treatment-seeking population - a cross-sectional study. Journal of Affective Disorders, 278:357-364.
DOI: https://doi.org/10.1016/j.jad.2020.08.074

Journal of Affective Disorders 278 (2021) 357–364

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Sex differences in depressive symptoms and their networks in a treatment-seeking population – a cross-sectional study

Johannes Simon Vetter[a,1,*], Tobias Raphael Spiller[b,1], Flurin Cathomas[a,c], Donald Robinaugh[d], Annette Brühl[a], Heinz Boeker[a], Erich Seifritz[a], Birgit Kleim[a]

[a] Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric Hospital, University of Zurich, Zurich, Switzerland
[b] Department of Consultation-Liaison Psychiatry and Psychosomatic Medicine, University Hospital Zurich, University of Zurich, Zurich, Switzerland
[c] Department of Neuroscience, Centre for Affective Neuroscience, and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, United States
[d] Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA, United States

ARTICLE INFO

Keywords:
Depression
Network analysis
Symptoms
Sex
Differences

ABSTRACT

Background: The higher prevalence of major depressive disorder (MDD) in females relative to males is well-established. Some authors have posited this difference arises to divergent symptom profiles in females vs. males. However, empirical tests of this hypothesis have yielded equivocal results. Here, we investigate sex differences in MDD of individual symptoms and symptom networks in a treatment-seeking sample.

Methods: We assessed depressive symptoms using Hamilton Depression Rating Scale (HDRS-17) in 590 treatment-seeking adults with MDD (300 females). We examined group differences in symptom endorsement. We investigated symptom networks and estimated Gaussian Graphical Models. Finally, we compared the female and male networks using the *Network Comparison Test*.

Results: Females scored significantly higher in psychological anxiety ($p < 0.001$; rB = -0.155), somatic anxiety ($p = .001$; rB = -0.150) and feelings of guilt ($p = .002$; rB = -0.139). Male and female patients did not differ in depression sum scores. There were no sex differences in network structure or global strength.

Limitations: Our study was sufficiently powered to detect only medium sized symptom differences. The generalizability of our study is limited to clinical samples and further studies are needed to investigate if findings also translate to outpatient samples.

Conclusion: Females reported elevated anxiety symptoms and guilt. Clinicians should assess these symptom

Financial support
JSV: None.
TRS: Forschungskredit of the University of Zurich, grant no. [FK-19–048].
FC: Early Postdoc Mobility Fellowship of the Swiss National Science Foundation and a Walter and Gertrud Siegenthaler Postdoctoral Fellowship (to FC).
DR.: DR's work was supported by funding from the NIMH (1K23MH113805–01A1).
AB: None.
HB: None.
ES: University of Zurich.
BK: None.

Conflicts of Interest:
JSV: None.
TRS: None.
FC: None.
DR.: None.
AB: None.
HB: None.
ES: None.
BK: None.

Acknowledgements
We would like to thank all patients as well as the physicians and psychologists for their help with obtaining data.

* Corresponding author: Psychiatric University Hospital, University of Zurich, Lenggstrasse 31, CH-8032 Zurich, Switzerland.
E-mail address: johannes.vetter@pukzh.ch (J.S. Vetter).

[1] These authors contributed equally to this paper

https://doi.org/10.1016/j.jad.2020.08.074
Received 9 February 2020; Received in revised form 20 July 2020; Accepted 25 August 2020
Available online 01 September 2020

differences and tailor treatment to individual symptom profiles. No differences between sexes emerged in MDD network structures, indicating that features may be more similar than previously assumed. Sex differences in psychopathological features of MDD are important for future research and personalized treatment.

1. Introduction

The increased prevalence of major depressive disorder (MDD) in females compared to males has been reported consistently over time and in different populations (Kessler et al., 2003; Lim et al., 2018). The reason for these differences in prevalence has been attributed to various underlying sex differences, for example in neurobiology (Rubinow and Schmidt, 2019) and the prevalence of subtypes of depression (Silverstein et al., 2017). With regard to the latter, multiple studies found that females are more often affected by "atypical" depression (Angst et al., 2002; Marcus et al., 2008). Atypical depression is defined by a specifier of MDD in different versions of the Diagnostic and Statistical Manual of Mental Disorders (DSM) by the predominance of symptoms such as weight gain or hypersomnia (Blanco et al., 2012; Marcus et al., 2008) compared to symptoms considered more typical such as disturbances of appetite and sleep disturbances. Depressed females tend to report more somatic symptoms (e.g., gastrointestinal symptoms and loss of appetite; (Angst et al., 2002; Silverstein et al., 2013) and further symptoms not included in the DSM definition of MDD (e.g., loss of libido and anxiety; Martin et al., 2013). One meta-analysis included data from 32 studies with a total of 108,260 individuals and found differences in the symptom profiles of females and males. Females reported higher intensity of somatic symptoms (e.g., sleep disturbance or fatigue) but also of core symptoms of depression (e.g., depressed mood or diminished interest; Cavanagh et al., 2017). Males presented more often with impulse control problems and substance use (e.g., nicotine/tobacco; Lamers et al., 2011). These differences, however, were moderate and heterogeneity was generally large. This might be due to the vast variety of measures used to assess depressive symptoms across studies, variances in definitions of depression and symptoms, and differences in participant samples and study designs (Cavanagh et al., 2017). Thus, it is not surprising that numerous different factor structures of depression have been proposed (Pancheri et al., 2002). Consequently, there is also evidence that differential symptom profiles may be associated with distinct risk factors (Fried et al., 2014). Several risk factors (depression history, childhood stress, stressful life events, and sex) all predicted progression of depressive symptoms, they each had a specific association to individual depressive symptoms (Fried et al., 2014). Moreover, there is evidence that cognitive and memory biases are more strongly related to some, but not all, symptoms of depression (Beevers et al., 2019; Marchetti et al., 2018). Finally, treatment of depression does not affect all symptoms uniformly (Boschloo et al., 2019; Mullarkey et al., 2020). Taken together, this adds to the importance of symptom-level approaches in depression research.

Differences on the symptom level are of relevance when depression is investigated from a network approach perspective (Fried et al., 2017). This approach conceptualizes mental disorders based on a complex systems theory framework as the interaction of their symptoms (Borsboom and Cramer, 2013). For example, in depression, insomnia may lead to concentration problems. This may then negatively impact one's performance at work, which again exacerbates insomnia, because one ruminates about the low performance at work (Cramer et al., 2016). According to Borsboom (2017), current conceptualizations of mental disorders presume an underlying latent disease entity to be the common cause of the symptoms that reflect its presence. From this perspective, symptoms are conceptualized to be diagnostically equivalent and interchangeable (Cramer et al., 2010; Lux and Kendler, 2010) and can therefore be summed up to an overall score indicating the severity of a mental disorder (Fried, 2015; Fried and Nesse, 2015b). This seems not only clinically implausible, but also negates the potential relevance of differences in symptom profiles of depression in females and males. While traditional conceptualizations have their shortcomings, novel approaches may be an alternative. The network approach posits that symptom co-occurrence arises not from a common underlying cause, but from symptom-to-symptom interactions (Borsboom, 2017). From this network perspective, differences between groups in the severity of individual symptoms or the interactions among those symptoms are exceedingly important as they indicate the possibility of differences between those groups in the causes of depression. Nevertheless, it has been noted, that the contrast between the two interpretations of correlations among symptoms (reflecting an underlying disorder versus a causal network), is likely less dogmatic than often pictured (Bringmann and Eronen, 2018). Furthermore, there is no general test to investigate if the "true" model is a network or a factor model (Fried, 2020).

Although there are numerous studies that investigate depression from a network perspective (e.g., van Borkulo et al., 2015), including studies about the association of depressive symptoms with environmental and genetic risk factors (van Loo et al., 2018) and co-expression of symptoms of anxiety and depression (Beard et al., 2016), only one study has analyzed sex differences (Mullarkey et al., 2018). Mullarkey et al. (2018) found that the depressive symptom networks of 646 male and 744 female adolescents from the general population differed in one relationship between two symptoms, namely that the association between self-hatred and negative body image was stronger in females. In another non-peer-reviewed preprint, van Borkulo et al. (2017) reported no differences between the symptom networks of depressive symptoms of 351 male and 701 female adults of a clinical population. To our knowledge, no published study has analyzed both sex differences in depression symptom networks in adults, despite the importance of these analyses for our understanding of depression and the role of sex in its development. The main purpose of the study was an explorative assessment of sex differences in depressive symptom profiles and symptom networks in a clinical sample of adults with a diagnosis of MDD. We investigated the interaction of these symptoms in treatment seeking patients of a psychiatric hospital.

2. Methods

2.1. Participants and procedure

As part of the regular clinical documentation at admission to the department for affective disorders at the Psychiatric Hospital of the University of Zurich, Switzerland, inpatients and day-clinic patients undergo basic clinical assessment. Data were collected as part of the routine clinical care procedure and completely anonymized. No specific written informed consent was thus obtained.

We analyzed data from patients admitted during July 2007 to June 2018. For this study, inclusion criteria were: age between 18 and 70, clinical diagnosis of major depressive disorder (single episode or recurrent; ICD-10 code F32 or F33; World Health Organization, 1992) as the main diagnosis, and data was only included from the first admission. To ensure the same power for detecting differences for all symptoms, we excluded participants if they failed to provide rating for all HDRS-17 items. Among the 611 patients who met the inclusion criteria, this led to the exclusion of additional 21 participants, leaving a final sample of 590 patients (300 female and 290 male patients).

3. Measures

Sex was determined as a binary variable (female, male). Individual sex was obtained from the patient's electronic patient record, which is

documented in accordance with the sex on the official ID-card or passport.

Depressive symptoms were assessed using the clinician-administered Hamilton Depression Rating Scale (HDRS-17; Hamilton, 1960). Clinically experienced physicians and psychologists completed HDRS questionnaires during the first three days of admission. The intensity of some symptoms (e.g., depressed mood, suicidal ideation, decreased interest, or psychomotor retardation) was rated from 0 to 4 (0 = none/absent to 4 = most severe), whereas other symptoms' intensity (e.g., decreased appetite or general somatic symptoms) was rated from 0 to 2 (0 = none/absent to 2 = severe). The HDRS-17 has extensively been used in research on depression (Trajković et al., 2011). The reliability of the total depression symptom scale derived in our study (Cronbach's $\alpha = 0.74$) was comparable to other studies (Trajković et al., 2011).

3.1. Statistical analyses

First, we compared females and males on both their overall HDRS score as well as their endorsement of individual HDRS items. Due to evidence of a non-normal distribution of the data, we used non-parametric Mann-Whitney U tests. To control for the family-wise error rate, Bonferroni correction was applied. Given the current sample size and global alpha level of 0.05, post-hoc power to detect small, medium and large effects ($d = 0.2, 0.5, 0.8$; Cohen, 1988), was 0.68, 1.00 and 1.00, respectively (calculated with g*power; Faul et al., 2009). Second, we used network analysis to examine relationships among HDRS items. We modeled our network analyses after those completed by Fried et al. (2017), and structured our results using the four components identified by those authors: (a) network estimation; (b) network characterization; (c) network stability, and (d) network comparison. All network analyses were carried out using R version 3.6.2. Networks were visualized using the R-package *qgraph* (Epskamp et al., 2012). All other statistical analyses were conducted with JASP Version 0.9.2.0 (JASP Team, 2020).

Given that a lack of variability of an item can bias the network structure (Terluin et al., 2016), we checked the variance of all items. One item (*Insight*) showed very low variability and was therefore excluded from the network analysis. In addition, prior to network estimation, we tested if the included symptoms overlapped, using the suggested default settings of the goldbricker function of the *networktools* package (Jones, 2019). This analysis suggested that the items *Insomnia – Early* and *Insomnia – Middle* had more than 75% (but less than 90%) of topologically overlapping correlations. Thus, we investigated the effect of removing each item individually on our results in two sensitivity analyses. Further details are available in the supplement.

3.1.1. Network estimation

Following the recommendations for network analysis with cross-sectional ordinal data (Costantini et al., 2017; Epskamp et al., 2018), we estimated partial correlation networks. In the resulting network (a Gaussian Graphical Model; GGM), nodes represent symptoms of depression, and edges represent partial correlations between symptoms. The techniques used to estimate the networks are both based on graphical lasso, which is a regularization procedure (Costantini et al., 2017; Epskamp et al., 2018). First, we used the Fused Graphical Lasso (FGL) to estimate the networks for females and males jointly. We used the information criterion based FGL, to facilitate comparison with individually estimated networks (Costantini et al., 2017). Second, we estimated the networks for the subsamples individually for technical reasons (e.g., stability analysis is only available for individually estimated networks, see below). We used the R-packages *bootnet* (Version 1.2.2) (Epskamp et al., 2018) and *EstimateGroupNetwork* (Costantini and Epskamp, 2017).

Table 1
Demographic and clinical characteristics ($n = 590$).

Variables	Females ($n = 300$)		Males ($n = 290$)		χ^2	*p*-value
	n	%	*n*	%		
MDD, single episode (F32)	120	40	113	39	.66	.797
MDD, recurrent episode (F33)	180	60	177	61	.66	.797
Medication[a]	219	73	213	73.4	.015	.902
Non-psychotropic drugs	88	29.3	109	37.6	4.52	.034
Antidepressants	198	66	190	65.5	.015	.902
Anxiolytics	48	16	35	12.1	1.89	.170
Detoxication and withdrawal	1	.3	3	1	1.08	.299
Hypnotics	28	9.3	24	8.3	.21	.651
Neuroleptics	84	28	65	22.4	2.44	.118
Mood stabilizers	36	12	38	13.1	.164	.686
Stimulants	5	1.7	10	3.4	1.89	.169
Patients with comorbid diagnoses[b]	160	53.3	165	56.9	.76	.384
F00 - F09	0	0	0	0	-	-
F10 - F19	53	17.7	80	27.6	8.31	.004*
F20 - F29	2	0.7	4	1.4	.74	.388
F30 - F39	12	4	8	2.8	.69	.405
F40 - F48	74	24.7	76	26.2	.19	.668
F50 - F59	21	7		1	13.45	.000*
F60 - F69	44	14.7	22	7.6	7.44	.006
F70 - F79	0	0	1	.3	1.04	.309
F80 - F89	0	0	0	0	-	-
F90 - F98	9	3	14	4.8	1.32	.252

MDD = Major depressive disorder

[a] Patients can take more than one drug.

[b] Patients can have more than one comorbid diagnosis.

* $p < .05$. ** $p < .01$. – Adjusted p-value for medication: 0.00625 – Adjusted p-value for comorbidities: 0.005

F00-F09: Mental disorders due to known physiological conditions, F10-F19: Mental and behavioral disorders due to psychoactive substance use, F20-F29: Schizophrenia, schizotypal and delusional disorders, F30-F39: Mood [affective] disorders, F40-F48: Neurotic, stress-related and somatoform disorders, F50-F59: Behavioral syndromes associated with physiological disturbances and physical factors, F60-F69: Disorders of adult personality and behavior, F70-F79: Mental retardation, F80-F89: Disorders of psychological development, F90-F98: Behavioral and emotional disorders with onset usually occurring in childhood and adolescence.

F30 - F39: Mood [affective] disorders, F40 - F48: Neurotic, stress-related and somatoform disorders, F50- F59: Behavioral syndromes associated with physiological disturbances and physical factors, F60 - F69: Disorders of adult personality and behavior, F70 - F79: Mental retardation, F80-F89: Disorders of psychological development, F90-F98: Behavioral and emotional disorders with onset usually occurring in childhood and adolescence.

We used the graphical lasso estimation procedure to estimate the networks (Epskamp et al., 2018). First, we used the Fused Graphical Lasso (FGL) to estimate the networks for females and males jointly. We used the information criterion based FGL, to facilitate comparison with individually estimated networks (Costantini et al., 2017). Second, we estimated the networks for the subsamples individually for technical reasons (e.g., stability analysis is only available for individually estimated networks, see below).

3.1.2. Network characterization

Subsequently, we examined the characteristics of individual nodes within the network. First, we computed undirected strength centrality to assess the relative interconnectedness of nodes in a network. Next, we estimated "predictability" using the *mgm* package (Haslbeck and Waldorp, 2016). "Predictability" is the upper bound of the shared variance of a given node (measured in R^2) with all its neighbors, assuming that all connections go towards the given node (Haslbeck and Fried, 2017).

3.1.3. Network stability

Stability and accuracy analyses, as well as the estimation of predictability are not yet available for jointly estimated networks. We used the R-package *bootnet* Version (1.1) (Epskamp et al., 2018) to assess the stability for the individually estimated networks, reflecting a lower bound for stability for the jointly estimated networks. *Bootnet* uses bootstrapping procedures to compute 95% confidence intervals for the edge weights, and to calculate the correlation-stability coefficient. We performed edge weights difference test and centrality difference tests. The results of all these analyses are outlined in the Supplementary materials.

3.1.4. Network comparison

We used the R-package *NetworkComparisonTest* (NCT; van Borkulo et al., 2017) to test for differences in network structure (assuming that the structure of both networks is exactly the same), global strength (assuming that overall connectivity in both networks is exactly the same) and edge strength (assuming that all edges of both networks are of similar strength) between the female and male symptom networks (van Borkulo et al., 2017).

4. Results

4.1. Sample characteristics

Average age (in years) did not differ between females (M = 43.44, SD = 12.80) and males (M = 43.99, SD = 12.26); t(1) = 0.524, p = .601). Female and male patients did not differ in their main diagnosis (F32 or F33), use of medication, or the comorbidity with an additional mental disorder. However, males were more likely to have an additional diagnosis of Mental and behavioral disorders due to psychoactive substance use (F10-F19). Females were more likely to have an additional diagnosis of disorders associated with physiological disturbances and physical factors (F55-F59). Further sample characteristics are presented in Table 1.

4.2. Individual MDD symptoms

We found no sex difference in the sum score of the HDRS-17. Regarding individual items, female patients had higher ratings in *Anxiety Psychic* ($p < 0.001$; W = 36,763; $r_B = -0.155$), *Anxiety Somatic* (p = .001; W = 36,979; $r_B = -0.150$) and *Feelings of Guilt* (p = .002; W = 37,468; $r_B = -0.139$). No item was more frequently endorsed by males. Mean scores and standard deviations for each of the 17 depression symptoms indexed by the HDRS are presented in Fig. 1. Two sensitivity analyses, individually excluding patients with an additional F10-F19 or F50-F55 diagnoses showed the same differences (for more details see the Supplementary materials).

4.3. MDD symptom networks

The networks for males and females are presented in Fig. 2. The NCT revealed no differences between sexes in network structure (p = .5578), or global strength (p = .2626; see R script for more details). Because the network structure was found to be invariant, we did not test individual connection strengths (van Borkulo et al., 2017). Neither removing *Insomnia – Middle* or *Insomnia – Early* from the network, as suggested by the goldbricker function, nor excluding patients with an additional F10-F19 or F50-F59 diagnosis altered the results (see Supplementary Materials). In accordance with the NCT, visual inspection of the network suggests that depressive symptom networks for females and males shared many edges and network features. Across both sexes, *Somatic Symptoms – Gastro-intestinal* and *Loss of Weight* were strongly connected. There were also strong connections among the three *Insomnia* items, as well as connections between *Anxiety Somatic* and *Hypochondriasis* in both males and females. *Agitation* showed the weakest edges in both sexes' networks.

4.4. Individual networks

4.4.1. Centrality

Node strength estimations are shown in Figure S1. For both sexes, nodes with high strength values were *Depressed Mood* and *Anxiety Somatic*. For females, the node with the highest strength was *Insomnia – Middle*, for males *Depressed Mood. Agitation* had lowest centrality estimates in both sexes. However, the CS-coefficient for strength for the female networks was 0.440 and the male network 0.438, both not exceeding the recommended threshold of 0.5 (Epskamp et al., 2018). In line with these results, the significance testing revealed that only *Agitation* had lower centrality than most other items in the female and male network (Figure S2 for females and Figure S3 for males). In the male

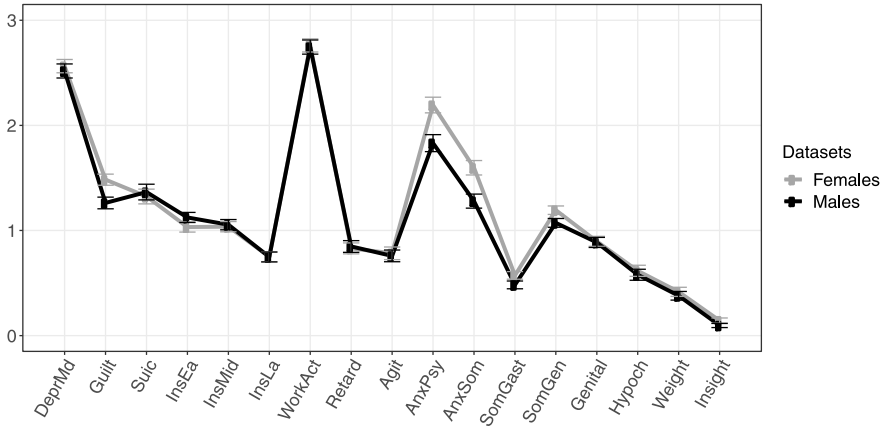

Fig. 1. Mean endorsement for each symptom for female (n = 300) and male patients (n = 290).
Standard error is presented in the figure by the error bars.
DeprMD = Depressed Mood, Guilt = Feelings of Guilt, Suic. = Suicidal ideation, InsEa = Insomnia – Early, InsMid = Insomnia – Middle, InsLa = Insomnia – Late, WorkAct = Work and Activities, Retard = Retardation, Agit = Agitation, AnxPsy = Anxiety Psychic, AnxSom = Anxiety Somatic, SomGast = Somatic Symptoms – Gastro-intestinal, SomGen = Somatic Symptoms – General, Genital = Genital Symptoms, Hypoch = Hypochondriasis, Weight = Loss of Weight; Insight = Insight.

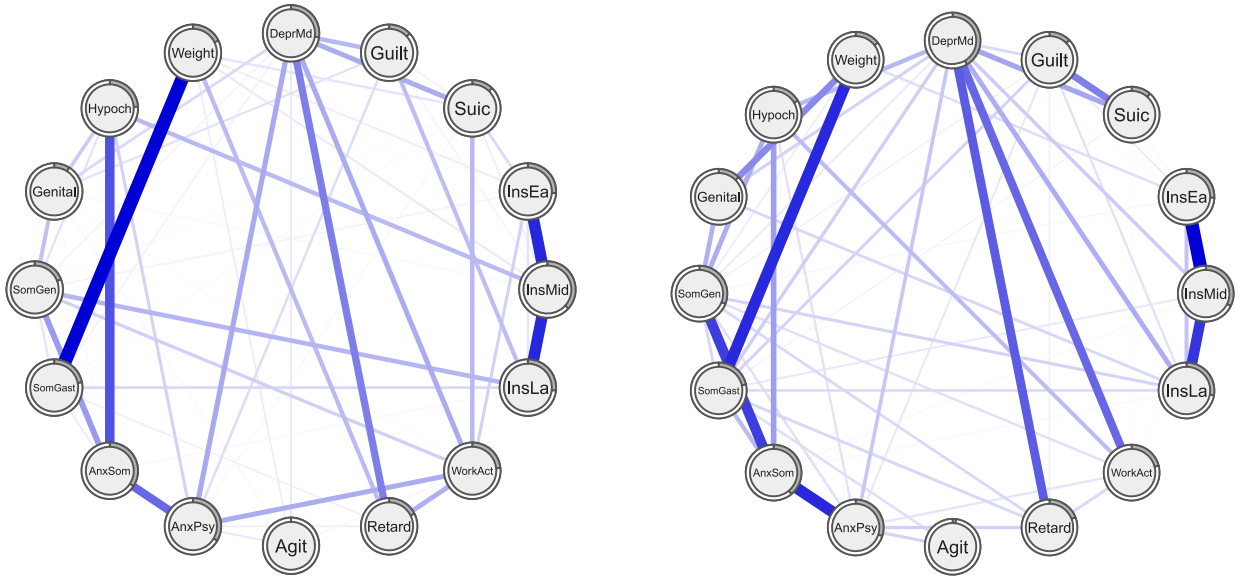

Fig. 2. Network of depressive symptoms for female (left, $n = 300$) and male (right, $n = 290$) treatment-seeking adults.
Nodes represent symptoms, edges partial correlations. Blue edges indicate positive, red edges negative correlations. The bigger the partial correlation, the ticker the edge. The gray area in the rings around the nodes depicts predictability (the upper bound of variance of a given node explained by all its neighbors), with a full circle corresponding to an R2 of 1.
DeprMD = Depressed Mood, Guilt = Feelings of Guilt, Suic. = Suicidal ideation, InsEa = Insomnia – Early, InsMid = Insomnia – Middle, InsLa = Insomnia – Late, WorkAct = Work and Activities, Retard = Retardation, Agit = Agitation, AnxPsy = Anxiety Psychic, AnxSom = Anxiety Somatic, SomGast = Somatic Symptoms – Gastro-intestinal, SomGen = Somatic Symptoms – General, Genital = Genital Symptoms, Hypoch = Hypochondriasis, Weight = Loss of Weight; Insight = Insight. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

network, *Depressed Mood* exhibited greater node strength than the majority of the other symptoms (Figure S3). Details and results of the significance tests for the individual edge weights are available in the Supplementary materials.

4.4.2. Predictability

For both sexes, *Agitation* had the lowest predictability. Highest predictability was found for *Insomnia – Middle* (0.367) in females and for *Depressed Mood* (0.414) in males. Mean predictability was 0.214 for females and 0.224 for males.

5. Discussion

The current study investigated sex differences in depressive symptom profiles and in the corresponding symptom network structure in an adult, treatment-seeking population suffering from major depressive disorder. Females reported more anxiety and more feelings of guilt than males. We did not find any additional sex differences in individual symptom endorsement or in the overall sum score. Moreover, no differences in symptom networks emerged.

In accordance with a meta-analysis on sex differences in depressive symptoms by Cavanagh et al. (2017), we found that females reported more anxiety than males. Moreover, this was the case for both psychic (e.g., disproportional worries or fear) and somatic symptoms (e.g., tremor or sweating) of anxiety. It is important to note that we based our analyses on a treatment seeking population and our data facilitated an estimation of the effect of comorbidities, while in the aforementioned meta-analysis (Cavanagh et al., 2017) none of the included studies reported any existing comorbidities. Also, in line with the meta-analysis was the small effect size for both differences. One reason females report more anxiety might be a higher prevalence of comorbid anxiety disorders in females, which has been well documented (McLean et al., 2011; Simonds and Whiffen, 2003). In our sample, however, the rate of comorbidity with a neurotic, stress-related, or somatoform disorders (F4; World Health Organization, 1992), which include anxiety disorders, was equal in both sexes, suggesting that differences in comorbid anxiety cannot account for our findings. Hence, further research is needed to identify and evaluate potential causes for the increased level of anxiety in females with MDD. Independent of the underlying cause, increased levels of anxiety have been shown to negatively impact the efficacy of treatment in MDD (Fava et al., 2008). Therefore, clinicians should carefully assess symptoms of anxiety in MDD patients, especially in females. Additionally, clinicians should be aware that males may tend to underreport anxiety symptoms (Bekker and van Mens-Verhulst, 2007), resulting in a reporting bias (but also see McLean and Hope, 2010).

We also found females to express more feelings of guilt than males. Comparing our result with the meta-analysis by Cavanagh et al. (2017) has limited validity, because they combined worthlessness and guilt in one item. Still, they did not find a significant difference for the composite item. Our results and also the effect sizes, however, are in line with the sex difference in experiencing guilt in general (Else-Quest et al., 2012). Because our sample consisted of treatment-seeking individuals, an additional reason for higher levels of guilt in females might be the increased attention to emotions in females with severe depression (Thayer et al., 2003). Hence, we suggest clinicians to be aware of elevated levels of guilt in this sub-population. Moreover, clinicians should be mindful of the presence of anxiety and guilt in female patients, considering the potential role of these symptoms within the patients' case conceptualizations and, therefore, assess these symptoms pro-actively. Our results may suggest a focus on anxiety and guilt as possible treatment targets.

No other sex differences in symptom severity emerged. This is also partially consistent with the meta-analysis by Cavanagh et al. (2017). They did not find sex differences with regard to suicidality, psychomotor retardation and hypochondriasis. However, in contrast with our findings, they reported that females suffered more from depressed mood, problems with sleep, somatic difficulties and problems with appetite or weight. There are several potential reasons for these different findings. First, the meta-analysis covered symptoms assessed with different questionnaires. Thus, the HDRS symptoms did not exactly match with the items analyzed in the meta-analysis. Second, regarding the results of the individual studies, the heterogeneity was mostly high (Cavanagh et al., 2017). This limits the grade of evidence of

the meta-analysis. One of multiple reasons underlying this heterogeneity is the varying method of symptom assessment. Several studies found significant differences in overall self-reported symptom severity. However, no significant sex differences were found in clinical assessments (Dekker et al., 2008; Kornstein et al., 2000; Scheibe et al., 2003). We used the same type of assessment, i.e., clinician-based, and this has to be considered. Third, the effect sizes reported in the meta-analysis by Cavanagh et al. (2017) were small or even very small. Our study achieved a post-hoc calculated power of 0.69 for small effect sizes after the correction for the family-wise error rate. Hence, our limited power may have resulted in false negative findings (see limitations).

We did not find any sex differences regarding the network structure or global network strength. This is in line with the results of Mullarkey et al. (2018) who reported no difference in global strength in depressive symptom networks of female and male adolescents. Although they found the network structure to be different across the two sexes, this was due to one out of dozens of edges. Furthermore, this edge was between two symptoms not commonly attributed to depression, namely self-hatred and negative body image. Our findings are also in line with a preprint by van Borkulo et al. (2017), which found no sex differences in symptom networks in depressed adults. There are several potential reasons for these null findings. First, the NCT is a conservative test (van Borkulo et al., 2017; Williams et al., 2019). Therefore, investigations with larger samples might be able to detect sex differences in network structures. Moreover, additional statistical tools for comparing networks are under current development (e.g., Haslbeck et al., 2019) and at least one will have higher power to find differences (Williams et al., 2019). Applying these methods, once available, might reveal sex differences. Second, statistical power to detect the effect of a variable on network structure depends on the sample and effect size, but also on the model (i.e., the symptoms included). A GGM is estimated on the basis of all individual symptoms' variance accounted for by all other included symptoms (Epskamp et al., 2018). Consequentially, when networks are compared with each other, only the shares of the variance accounted for by the networks are compared with each other (e.g., using the NCT; van Borkulo et al., 2017). Hence, given an effect and sample size, the statistical power of the NCT directly relates to the total amount of variance accounted for by the models which are compared with each other. Predictability is an estimate for the upper bound of variance (measured in R^2) of a given symptom explained for by all other items in the network. In our study, mean predictability was approximately 0.2 for both sexes, which is comparable to the mean predictability in other studies of networks of depressive symptoms (Haslbeck and Fried, 2017). In other words, on average at least 80% of a symptom's variance was not explained by other symptoms. This additionally limits the power of the NCT. Moreover, such values for mean predictability are not only low absolute values, but also relatively low, when compared to mean predictability in network models of other disorders (Haslbeck and Fried, 2017). This indicates that symptom network models of depression perform badly at capturing the variance in symptoms of depression. This poses a significant challenge for all network models of MDD. One solution is comparing predictability of network models based on different symptoms sampled out of a set of depressive symptoms and selecting a model with high predictability. However, individual symptom profiles of MDD are known to have substantial heterogeneity (Fried et al., 2016; Fried and Nesse, 2015a), which might limit the optimization of predictability in group-level network models of MDD per se. Third, as outlined above, sex differences in symptom profiles have a small effect size and individual findings are heterogeneous (Cavanagh et al., 2017). Furthermore, given the high heterogeneity of individual symptom profiles (Fried and Nesse, 2015a), one could question the utility of group level network models of symptoms of depression at all. Moreover, the network approach perspective is primarily focused on within-subject, rather than between-subject effects (Borsboom, 2017). Thus, investigations of network models based on within-subject effects (e.g., Bringmann et al., 2015) may provide important additional information about the relationships among depression symptoms in males and females.

5.1. Limitations

This study has several limitations. First, our analysis focused on individual HDRS items and we did not investigate sex differences on the factor level of the HDRS (e.g., based on the factor structure described by Cleary and Guy (1977). Due to our focus on differences on the level of individual HDRS items, our limited sample size and the fact that the software to conduct latent network analysis (Epskamp et al., 2017b) was still under development, we did not estimate a latent network. Nevertheless, further research is needed to explore potential sex differences in latent networks of depression. Furthermore, such an investigation could also address a second limitation of our study, namely that some of the symptoms assessed with the HDRS probably measure overlapping constructs, which artificially inflates edge weights between these items. This was confirmed by the goldbricker function indicating topologically overlap between the items *Insomnia – Early* and *Insomnia – Middle*. As noted above, however, networks of females and males did not differ when either one of these items were excluded from the analysis. Third, as outlined above, our post-hoc calculated power to detected differences in symptom profiles with small effect size was 0.69. It is possible that we would have found more sex differences with greater power. However, to detect small effect sizes with sufficient power (0.9), while also accounting for the family-wise error rate, an approximate sample size of 1830 would have been needed. Fourth, although the HDRS is one of the most commonly used rating scales for depression (Santor et al., 2006), it does not reflect the current DSM-5 or ICD-10 definition of MDD. Still, its broad use enables comparability with existing literature. Fifth, the investigation of a clinical compared to a healthy population leads to spurious negative or weaker correlations between symptoms in a network due to Berkson's bias (de Ron et al., 2019). This limits the generalizability of our findings based on network analysis to non-clinical populations. Sixth, to our knowledge, there is no possibility to calculate or estimate post-hoc power of an NCT-analysis. Thus, effective power of our analyses remains undetermined. In addition, the NCT is not designed to collect evidence for the null hypothesis. Consequently, we could not assess if the null hypothesis (that there are no sex differences in network structures) is true. Nevertheless, with future methodological advances (e.g., Bayesian methods for estimating GGMs; Williams et al., 2019) this will be possible. Lastly, sex was assessed as a binary variable based on participants' sex in their official identity documents. Transgender people in Switzerland can change their sex in their identity documents but must choose between male or female. Therefore, no third option for sex besides female or male was available.

6. Future directions

Based on our results and the limitations of our study, we have several suggestions for future research. With regard to sex differences of depressive symptom profiles, the heterogeneity of the existing literature should be addressed systematically. As an example, there is evidence that females with severe depression are more aware of their emotions (Thayer et al., 2003). Consequentially, the kind of symptom assessment (as a self-report or by a clinician) likely influences symptom reporting and comparison of symptom severity between females and males. We have several additional recommendations for research on sex differences in network models of MDD to the ones outlined above. First, our study needs to be replicated, especially with a bigger sample size to increase power. Second, future studies should investigate network structure in community samples, which avoids the induction of Berkson's bias. However, these results could only be compared to clinical samples under the assumption that the underlying network structure is independent of the severity of the symptoms. Third, we strongly recommend future studies to estimate predictability to assess the limits of their model in explaining individual symptoms' variance. Our findings provide evidence, however, that there are no, or only subtle, sex differences in the network structure of depressive symptoms.

7. Conclusion

This study is amongst the first to investigate sex differences in depressive symptoms and their corresponding network structure in adults. We found that females reported more symptoms of anxiety and guilt, which is in line with the literature. However, in a recent meta-analysis (Cavanagh et al., 2017), females were not found to have more symptoms of guilt. Similar to the two existing similar studies (Mullarkey et al., 2018; van Borkulo et al., 2017), we found no sex differences in symptom network structures. Given that the sex differences in individual depressive symptoms were of small effect, one potential reason for our null finding is the design of the used network analytic method, which is aimed to optimize specificity at cost of lower sensitivity. Taken together, our results indicate, that sex differences in depressive symptoms in treatment-seeking adults are few and subtle.

Author statement

All authors state that research was conducted in accordance with the Helsinki Declaration as revised 1989.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jad.2020.08.074.

References

Angst, J., Gamma, A., Gastpar, M., Lepine, J.-.P., Mendlewicz, J., Tylee, A., Depression Research in European Society Study, 2002. Gender differences in depression. Epidemiological findings from the European DEPRES I and II studies. Eur. Arch. Psychiatry Clin. Neurosci. 252 (5), 201–209. https://doi.org/10.1007/s00406-002-0381-6.

Beard, C., Millner, A.J., Forgeard, M.J.C., Fried, E.I., Hsu, K.J., Treadway, M.T., Leonard, C.V., Kertz, S.J., Björgvinsson, T., 2016. Network analysis of depression and anxiety symptom relationships in a psychiatric sample. Psychol. Med. 46 (16), 3359–3369. https://doi.org/10.1017/S0033291716002300.

Beevers, C.G., Mullarkey, M.C., Dainer-Best, J., Stewart, R.A., Labrada, J., Allen, J.J.B., McGeary, J.E., Shumake, J., 2019. Association between negative cognitive bias and depression: a symptom-level approach. J. Abnorm. Psychol. 128 (3), 212–227. https://doi.org/10.1037/abn0000405.

Bekker, M.H.J., van Mens-Verhulst, J., 2007. Anxiety disorders: sex differences in prevalence, degree, and background, but gender-neutral treatment. Gend. Med. 4 (Suppl B), S178–S193.

Blanco, C., Vesga-López, O., Stewart, J.W., Liu, S.-.M., Grant, B.F., Hasin, D.S., 2012. Epidemiology of major depression with atypical features: results from the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). J. Clin. Psychiatry 73 (2), 224–232. PubMed.. https://doi.org/10.4088/JCP.10m06227.

Borsboom, D., 2017. A network theory of mental disorders. World Psychiatry 16 (1), 5–13. https://doi.org/10.1002/wps.20375.

Borsboom, D., Cramer, A.O.J., 2013. Network Analysis: an Integrative Approach to the Structure of Psychopathology. Annu. Rev. Clin. Psychol. 9 (1), 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608.

Boschloo, L., Cuijpers, P., Karyotaki, E., Berger, T., Moritz, S., Meyer, B., Klein, J.P., 2019. Symptom-specific effectiveness of an internet-based intervention in the treatment of mild to moderate depressive symptomatology: the potential of network estimation techniques. Behav. Res. Ther. 122, 103440. https://doi.org/10.1016/j.brat.2019.103440.

Bringmann, L.F., Eronen, M.I., 2018. Don't blame the model: reconsidering the network approach to psychopathology. Psychol. Rev. 125 (4), 606–615. https://doi.org/10.1037/rev0000108.

Bringmann, L.F., Lemmens, L.H.J.M., Huibers, M.J.H., Borsboom, D., Tuerlinckx, F., 2015. Revealing the dynamic network structure of the Beck Depression Inventory-II. Psychol. Med. 45 (4), 747–757. Cambridge Core.. https://doi.org/10.1017/S0033291714001809.

Cavanagh, A., Wilson, C.J., Kavanagh, D.J., Caputi, P., 2017. Differences in the expression of symptoms in men versus women with depression: a systematic review and meta-analysis. Harv. Rev. Psychiatry 25 (1), 29–38. https://doi.org/10.1097/HRP.0000000000000128.

Cleary, P., Guy, W., 1977. Factor analysis of the Hamilton depression scale. Drugs Exp. Clin. Res. 1 (1–2), 115–120.

Cohen, J., 1988. Statistical Power Analysis For the Behavioral Sciences, 2nd ed. L. Erlbaum Associates.

Costantini, G., Epskamp, S., 2017. EstimateGroupNetwork: Perform the Joint Graphical Lasso and Selects Tuning Parameters. *R package (Version 0.1.2)*.

Costantini, G., Richetin, J., Preti, E., Casini, E., Epskamp, S., Perugini, M., 2017. Stability and variability of personality networks. A tutorial on recent developments in network psychometrics. Pers. Individ. Dif. https://doi.org/10.1016/j.paid.2017.06.011.

Cramer, A.O.J., van Borkulo, C.D., Giltay, E.J., van der Maas, H.L.J., Kendler, K.S., Scheffer, M., Borsboom, D., 2016. Major Depression as a Complex Dynamic System. PLoS ONE 11 (12), e0167490. https://doi.org/10.1371/journal.pone.0167490.

Cramer, A.O.J., Waldorp, L.J., van der Maas, H.L.J., Borsboom, D., 2010. Comorbidity: a network perspective. Behav. Brain Sci. 33 (2–3), 137–150. https://doi.org/10.1017/S0140525X09991567.

de Ron, J., Fried, E.I., Epskamp, S., 2019. *Psychological Networks in Clinical Populations: A tutorial On the Consequences of Berkson's Bias* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/5t8zw.

Dekker, J., Koelen, J.A., Peen, J., Schoevers, R.A., Gijsbers-Van Wijk, C., 2008. Gender differences in clinical features of depressed outpatients: preliminary evidence for subtyping of depression? Women Health 46 (4), 19–38. https://doi.org/10.1300/J013v46n04_02.

Else-Quest, N.M., Higgins, A., Allison, C., Morton, L.C., 2012. Gender differences in self-conscious emotional experience: a meta-analysis. Psychol. Bull. 138 (5), 947–981. https://doi.org/10.1037/a0027930.

Epskamp, S., Borsboom, D., Fried, E.I., 2018. Estimating psychological networks and their accuracy: a tutorial paper. Behav. Res. Methods 50 (1), 195–212. https://doi.org/10.3758/s13428-017-0862-1.

Epskamp, S., Cramer, A.O.J., Waldorp, L.J., Schmittmann, V.D., Borsboom, D., 2012. qgraph: network visualizations of relationships in psychometric data. J. Stat. Softw. 48 (4), 1–18. https://doi.org/10.18637/jss.v048.i04.

Epskamp, S., Rhemtulla, M., Borsboom, D., 2017b. Generalized network psychometrics: combining network and latent variable models. Psychometrika 82 (4), 904–927. https://doi.org/10.1007/s11336-017-9557-x.

Faul, F., Erdfelder, E., Buchner, A., Lang, A.-.G., 2009. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav. Res. Methods 41 (4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149.

Fava, M., Rush, A.J., Alpert, J.E., Balasubramani, G.K., Wisniewski, S.R., Carmin, C.N., Biggs, M.M., Zisook, S., Leuchter, A., Howland, R., Warden, D., Trivedi, M.H., 2008. Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. Am. J. Psychiatry 165 (3), 342–351. https://doi.org/10.1176/appi.ajp.2007.06111868.

Fried, E.I., 2015. Problematic assumptions have slowed down depression research: why symptoms, not syndromes are the way forward. Front. Psychol. 6. https://doi.org/10.3389/fpsyg.2015.00309.

Fried, E.I., 2020. *Lack of Theory Building and Testing Impedes Progress in the Factor and Network Literature* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/zg84s.

Fried, E.I., Eidhof, M., Palic, S., Costantini, G., H.-v. Dijk, Hilde, O., Claudi L.H., Engelhard, I., Armour, C., Nielsen, A.B., & Karstoft, K.-I.. (2017). *Replicability and generalizability of PTSD networks: A cross-cultural multisite study of PTSD symptoms in four trauma patient samples.* https://doi.org/10.17605/OSF.IO/3ZQ5U.

Fried, E.I., Nesse, R.M., 2015a. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR*D study. J. Affect. Disord. 172, 96–102. https://doi.org/10.1016/j.jad.2014.10.010.

Fried, E.I., Nesse, R.M., 2015b. Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. BMC Med. 13 (1). https://doi.org/10.1186/s12916-015-0325-4.

Fried, E.I., Nesse, R.M., Zivin, K., Guille, C., Sen, S., 2014. Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors. Psychol. Med. 44 (10), 2067–2076. https://doi.org/10.1017/S0033291713002900.

Fried, E.I., van Borkulo, C.D., Cramer, A.O.J., Boschloo, L., Schoevers, R.A., Borsboom, D., 2017b. Mental disorders as networks of problems: a review of recent insights. Soc. Psychiatry Psychiatr. Epidemiol. 52 (1), 1–10. https://doi.org/10.1007/s00127-016-1319-z.

Fried, E.I., van Borkulo, C.D., Epskamp, S., Schoevers, R.A., Tuerlinckx, F., Borsboom, D., 2016. Measuring depression over time . . . Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. Psychol. Assess. 28 (11), 1354–1367. https://doi.org/10.1037/pas0000275.

Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psychiatr. 23 (1), 56. http://dx.doi.org/10.1136/jnnp.23.1.56.

Haslbeck, J.M.B., Borsboom, D., Waldorp, L., 2019. Moderated network models. ArXiv:1807.02877 [Stat]. http://arxiv.org/abs/1807.02877.

Haslbeck, J.M.B., Fried, E.I., 2017. How predictable are symptoms in psychopathological networks? A reanalysis of 18 published datasets. Psychol. Med. 47 (16), 2767–2776. https://doi.org/10.1017/S0033291717001258.

Haslbeck, J.M.B., & Waldorp, L.J. (2016). *mgm- structure estimation for time-varying mixed graphical models in high-dimensional data.*

JASP Team. (2020). *JASP* (Version 0.12.2) [Computer Software]. https://jasp-stats.org.

Jones, P. (2019). *Networktools: tools for identifying important nodes in networks* (Version 1.2.1) [Computer Software]. https://CRAN.R-project.org/package=networktools.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K.R., Rush, A.J., Walters, E.E., Wang, P.S., National Comorbidity Survey Replication, 2003. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA 289 (23), 3095–3105. https://doi.org/10.1001/jama.289.23.3095.

Kornstein, S.G., Schatzberg, A.F., Thase, M.E., Yonkers, K.A., McCullough, J.P., Keitner, G.I., Gelenberg, A.J., Ryan, C.E., Hess, A.L., Harrison, W., Davis, S.M., Keller, M.B., 2000. Gender differences in chronic major and double depression. J. Affect. Disord. 60 (1), 1–11. https://doi.org/10.1016/S0165-0327(99)00158-5.

Lamers, F., van Oppen, P., Comijs, H.C., Smit, J.H., Spinhoven, P., van Balkom, A.J.L.M., Nolen, W.A., Zitman, F.G., Beekman, A.T.F., Penninx, B.W.J.H., 2011. Comorbidity patterns of anxiety and depressive disorders in a large cohort study: the Netherlands Study of Depression and Anxiety (NESDA). J. Clin. Psychiatry 72 (3), 341–348. https://doi.org/10.4088/JCP.10m06176blu.

Lim, G.Y., Tam, W.W., Lu, Y., Ho, C.S., Zhang, M.W., Ho, R.C., 2018. Prevalence of depression in the community from 30 countries between 1994 and 2014. Sci. Rep. 8 (1), 2861. https://doi.org/10.1038/s41598-018-21243-x.

Lux, V., Kendler, K.S., 2010. Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria. Psychol. Med. 40 (10), 1679–1690. https://doi.org/10.1017/S0033291709992157.

Marchetti, I., Everaert, J., Dainer-Best, J., Loeys, T., Beevers, C.G., Koster, E.H.W., 2018. Specificity and overlap of attention and memory biases in depression. J. Affect. Disord. 225, 404–412. https://doi.org/10.1016/j.jad.2017.08.037.

Marcus, S.M., Kerber, K.B., Rush, A.J., Wisniewski, S.R., Nierenberg, A., Balasubramani, G.K., Ritz, L., Kornstein, S., Young, E.A., Trivedi, M.H., 2008. Sex differences in depression symptoms in treatment-seeking adults: confirmatory analyses from the Sequenced Treatment Alternatives to Relieve Depression study. Compr. Psychiatry 49 (3), 238–246. PubMed.. https://doi.org/10.1016/j.comppsych.2007.06.012.

Martin, L.A., Neighbors, H.W., Griffith, D.M., 2013. The experience of symptoms of depression in men vs women: analysis of the national comorbidity survey replication. JAMA Psychiatry 70 (10), 1100. https://doi.org/10.1001/jamapsychiatry.2013.1985.

McLean, C.P., Asnaani, A., Litz, B.T., Hofmann, S.G., 2011. Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. J. Psychiatr. Res. 45 (8), 1027–1035. https://doi.org/10.1016/j.jpsychires.2011.03.006.

McLean, C.P., Hope, D.A., 2010. Subjective anxiety and behavioral avoidance: gender, gender role, and perceived confirmability of self-report. J. Anxiety Disord. 24 (5), 494–502. https://doi.org/10.1016/j.janxdis.2010.03.006.

Mullarkey, M.C., Marchetti, I., Beevers, C.G., 2018. Using network analysis to identify central symptoms of adolescent depression. J. Clin. Child Adoles. Psychol. 1–13. https://doi.org/10.1080/15374416.2018.1437735.

Mullarkey, M.C., Stein, A.T., Pearson, R., Beevers, C.G., 2020. Network analyses reveal which symptoms improve (or not) following an Internet intervention (Deprexis) for depression. Depress. Anxiety 37 (2), 115–124. https://doi.org/10.1002/da.22972.

Pancheri, P., Picardi, A., Pasquini, M., Gaetano, P., Biondi, M., 2002. Psychopathological dimensions of depression: a factor study of the 17-item Hamilton depression rating scale in unipolar depressed outpatients. J. Affect. Disord. 68 (1), 41–47. https://doi.org/10.1016/s0165-0327(00)00328-1.

Rubinow, D.R., Schmidt, P.J., 2019. Sex differences and the neurobiology of affective disorders. Neuropsychopharmacology 44 (1), 111. https://doi.org/10.1038/s41386-018-0148-z.

Santor, D.A., Gregus, M., Welch, A., 2006. Eight decades of measurement in depression. Measurement 4 (3), 135–155. https://doi.org/10.1207/s15366359mea0403_1.

Scheibe, S., Preuschhof, C., Cristi, C., Bagby, R.M., 2003. Are there gender differences in major depression and its response to antidepressants? J. Affect. Disord. 75 (3), 223–235. https://doi.org/10.1016/S0165-0327(02)00050-2.

Silverstein, B., Ajdacic-Gross, V., Rossler, W., Angst, J., 2017. The gender difference in depressive prevalence is due to high prevalence of somatic depression among women who do not have depressed relatives. J. Affect. Disord. 210, 269–272. https://doi.org/10.1016/j.jad.2017.01.006.

Silverstein, B., Edwards, T., Gamma, A., Ajdacic-Gross, V., Rossler, W., Angst, J., 2013. The role played by depression associated with somatic symptomatology in accounting for the gender difference in the prevalence of depression. Soc. Psychiatry Psychiatr. Epidemiol. 48 (2), 257–263. https://doi.org/10.1007/s00127-012-0540-7.

Simonds, V.M., Whiffen, V.E., 2003. Are gender differences in depression explained by gender differences in co-morbid anxiety? J. Affect. Disord. 77 (3), 197–202. https://doi.org/10.1016/S0165-0327(02)00113-1.

Terluin, B., Boer, M.R.de, Vet, H.C.W.de, 2016. Differences in connection strength between mental symptoms might be explained by differences in variance: reanalysis of network data did not confirm staging. PLoS ONE 11 (11), e0155205. https://doi.org/10.1371/journal.pone.0155205.

Thayer, J.F., Rossy, L.A., Ruiz-Padial, E., Johnsen, B.H., 2003. Gender differences in the relationship between emotional regulation and depressive symptoms. Cognit. Ther. Res. 27 (3), 349–364. https://doi.org/10.1023/A:1023922618287.

Trajković, G., Starčević, V., Latas, M., Leštarević, M., Ille, T., Bukumirić, Z., Marinković, J., 2011. Reliability of the Hamilton Rating Scale for depression: a meta-analysis over a period of 49years. Psychiatry Res. 189 (1), 1–9. https://doi.org/10.1016/j.psychres.2010.12.007.

van Borkulo, C.D., Boschloo, L., Borsboom, D., Penninx, B.W.J.H., Waldorp, L.J., Schoevers, R.A., 2015. Association of Symptom Network Structure With the Course of Depression. JAMA Psychiatry 72 (12), 1219. https://doi.org/10.1001/jamapsychiatry.2015.2079.

van Borkulo, C.D., Boschloo, L., Kossakowski, J., Tio, P., Schoevers, R., Borsboom, D., & Waldorp, L. (2017). *Comparing network structures on three aspects: a permutation test.* https://doi.org/10.13140/RG.2.2.29455.38569.

van Loo, H.M., Van Borkulo, C.D., Peterson, R.E., Fried, E.I., Aggen, S.H., Borsboom, D., Kendler, K.S., 2018. Robust symptom networks in recurrent major depression across different levels of genetic and environmental risk. J. Affect. Disord. 227, 313–322. https://doi.org/10.1016/j.jad.2017.10.038.

Williams, D.R., Rast, P., Pericchi, L., & Mulder, J. (2019). *Comparing Gaussian graphical models with the posterior predictive distribution and bayesian model selection* [preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/yt386.

World Health Organization, 1992. The ICD-10 Classification of Mental and Behavioural Disorders: Clinical descriptions and Diagnostic Guidelines. World Health Organization.